

Alternative Random Assignment Models

SOCIAL EXPERIMENTATION:
EVALUATING PUBLIC PROGRAMS WITH EXPERIMENTAL METHODS

PART 3

This is the third in a series of technical papers on social experiments. Previous papers discussed the rationale and background of social experimentation and presented some basic concepts and principles in the context of simple two-group experiments. This paper focuses on the design of more complex experiments to address a variety of policy questions. By **random assignment model**, we mean the way random assignment is integrated into the program intake process and the treatments to which individuals are assigned.

The way in which random assignment is integrated into the program intake process depends on the population of interest (*e.g.*, participants or program eligibles) and the nature of the program—*i.e.*, whether it is voluntary or mandatory and whether it is an ongoing program or a special demonstration set up specifically for experimental purposes. This integration is critical to the design of the experiment because it determines the composition of the experimental sample and, therefore, the population to which the experimental estimates will apply. The treatments to which individuals are assigned will depend on the policy question to be addressed.

We begin by describing how random assignment is typically integrated into the intake process in simple two-group experiments designed to address the question, does the program achieve its objectives? In the remainder of the paper, we consider random assignment models that address other questions.

Integrating Random Assignment into the Intake Process

In this section we examine the integration of random assignment into the program intake process in three different contexts: special demonstrations designed to estimate impacts on participants in voluntary programs, special demonstrations designed to estimate impacts on the entire population eligible for either voluntary or mandatory programs, and evaluations conducted in the context of ongoing programs. In a subsequent paper, we will discuss the implementation of these designs in the field. Here we focus on the implications of the integration of random assignment into its programmatic context for the composition of the experimental sample.

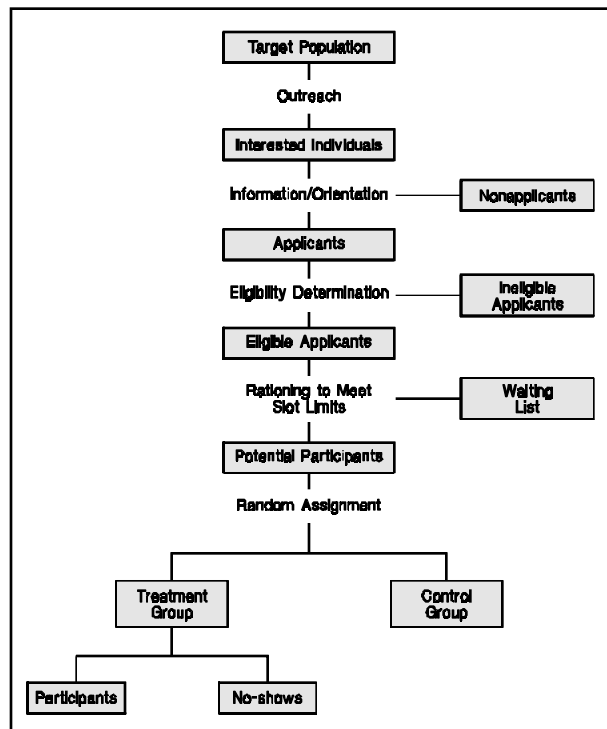
Special Demonstrations Designed to Estimate Impacts on Participants in Voluntary Programs

In voluntary programs, interest usually centers on the effects on program participants, on the presumption that nonparticipants are likely to be unaffected by the program. (This is not necessarily the case for mandatory programs, in which eligible individuals who do not participate face some penalty.) Exhibit 1 shows schematically the intake process, including random assignment, for a voluntary program (see next page).

As shown at the top of the exhibit, the intake process begins with program outreach to the target population in order to recruit applicants. Outreach can take any of a number of forms. For example, outreach for an employment program for welfare recipients might be conducted through telephone calls from case workers to selected recipients or notices sent with the monthly welfare check. Outreach for a home care program for the frail elderly might involve soliciting referrals from social agencies that serve the elderly, posting notices in adult day care centers, running ads in local newspapers, and/or announcements on television and radio.

Design for Estimating Impacts on Participants in a Voluntary Program

EXHIBIT 1



Individuals who respond to program outreach (the “interested individuals” in Exhibit 1) are given more detailed information about the program and the evaluation, to allow them to make a decision about whether to apply. This information may be provided through discussions with intake workers, either over the telephone or in person, written materials, or formal group orientation sessions. It includes a description of the services or benefits provided by the program, what program participation will entail for the individual, and the random assignment process and its implications for the individual, as well as any data collection procedures that will impinge on sample members. Upon receiving this information, some individuals formally apply for program entry (the “applicants”) while others decide not to (the “nonapplicants”).

Most social programs have some eligibility requirements. In some cases, there are relatively formal admission criteria, such as age, residence, family income, or receipt of welfare. For example, a community service program might be restricted to out-of-school youths between the ages of 18 and 25; a home care program for the elderly might be restricted to individuals over the age of 65 who live alone. Information on these objective applicant characteristics is typically collected on the program application form or in

documentation collected as part of the application. In other cases, eligibility criteria can be quite informal and judgmental, such as the intake worker’s assessment of the individual’s potential to benefit from the program. These judgments are usually formed on the basis of personal interviews with the applicant or with other social service providers who are familiar with the applicant (e.g., a social worker or health professional who has worked with the applicant). As shown in Exhibit 1, application of these criteria creates a set of “eligible applicants” and a set of “ineligible applicants,” who are excluded from further consideration.

In a regular program or a nonexperimental demonstration, if the number of eligible applicants exceeds the number of participant slots in the program, program staff must apply some further criteria to decide which will be admitted to the program. Since all eligible applicants are, by definition, eligible to participate, some criterion other than eligibility must be used. Eligible applicants might be admitted to the program on a first-come, first-served basis, with the remainder being placed on a waiting list. Or program staff might choose among eligible applicants on the basis of “deservingness” or need for program services. These choices produce a set of “potential participants” who are then offered admission to the program.

In an experiment, program staff use the same type of eligibility and selection criteria to choose a set of potential participants equal to the number of program slots to be filled plus the experimental control group. This group is then randomly assigned to one or more treatment groups, which are allowed to enter the program, or to a control group, which is excluded from the program. Under random assignment, both the treatment and control groups are representative of the pool of individuals from which they are drawn. Thus, *the experimental treatment and control groups represent the population of potential participants selected by program staff as being both eligible for the program and acceptable on the criteria (if any) intake staff would normally use to choose among eligible applicants when there are more than the program can accommodate.*

Not all of those randomly assigned to the treatment group will actually participate in the program, just as not all of the potential participants accepted by the program would have participated in the absence of the experiment. Some will lose interest or change their minds at the last minute, or other events in their lives (illness, a family crisis, finding a job on their own) will intervene and cause them to withdraw. These individuals are labeled “no-shows” in Exhibit 1; individuals who enter the program are labeled “participants”.

Several important features of this intake and random assignment process should be noted. First, up to the point of random assignment, the process departs from the intake process that would be employed in a regular program in only two ways. The first difference is that potential applicants must be informed about the experiment. The second, less obvious, difference is that in order to provide for a control group, the program must recruit more applicants than it normally would. Suppose, for example, that the program is designed to serve 90 participants and the expected no-show rate is 10 percent. In that case, 100 potential participants would have to be accepted into the program to fill the program's slots. If one-half of all eligible applicants are assigned to a control group, this means that intake staff must identify 200 potential participants in order to generate a treatment group of 100.

With these two exceptions, program participants are chosen exactly the same way they would be chosen in the absence of the experiment, *i.e.*, using the same processes and criteria that would be used in a nonexperimental program. Therefore, so long as these two differences do not change the nature of the pool of potential participants, the treatment group will accurately represent the group that would be offered admission to the program in the absence of an experiment.

A second point to be noted is that random assignment could have taken place at any point in the intake process. For example, interested individuals (the second box from the top of Exhibit 1), rather than potential participants, could have been randomly assigned to treatment and control groups. If that had been done, the treatment group would have gone on to the further steps in the intake process (application, eligibility determination, and selection of potential participants), while the control group would have been dropped from further consideration. This would have created comparable treatment and control groups composed of individuals who had expressed interest in the program and experimental impact estimates could have been derived for this group. In most cases, it would also be possible to derive experimental estimates of program impact on participants under this design, using the no-show adjustment described in the previous paper.¹ However, as we shall see in a subsequent paper, the precision of those estimates would be much lower than under a design that randomly assigned potential participants because of the larger number of nonparticipants in the interested individual population. As we shall see, when policy interest focuses on impacts on participants, it is preferable to position random assignment as late in the intake process as possible.

¹ It would also be possible to derive estimates of the impacts on applicants, eligible applicants, or potential participants, using variants of the no-show adjustment.

Special Demonstrations Designed to Estimate Impacts on the Eligible Population

In some cases, policy interest focuses on the impacts of the program on the entire population eligible for the program, not just those who participate. When the impact on nonparticipants can safely be assumed to be zero, and the size of the eligible population is known, the impact on the eligible population can be inferred from the impact on participants through an adjustment procedure analogous to the no-show adjustment. Suppose, for example, that 25 percent of those eligible for a training program participate and that the program raises the average annual earnings of participants by \$1,000. Then the average impact on the eligible population would be \$250 ($= 0.25 \times \$1,000$).

When the program may have nonzero impacts on nonparticipants, however, the experiment must be designed to measure impacts on the entire eligible population in order to capture all of its effects. This would be the case, for example, with mandatory programs. Consider, for instance, a mandatory work program for welfare recipients with youngest children over the age of 2. Not all welfare recipients in the mandatory population will participate in the program, either because they refuse or are exempted, or because the program may not be able to handle all those formally required to participate. But nonparticipants' behavior and circumstances may still be affected by the program. Those who refuse to participate may be sanctioned, affecting their welfare benefits or, possibly, increasing their willingness to find work on their own. Those whom the program has not yet required to participate may also be motivated to find work on their own before they are required to. Some recipients may find the requirement so onerous that they leave welfare altogether rather than participating. Even the behavior of those who are exempted from the requirement may be affected. In fact, recipients may change their behavior *in order to be exempted*—*e.g.*, by having another child or enrolling in an educational program. For all of these reasons, the assumption of zero impacts on nonparticipants may be untenable in mandatory programs.

Even in the case of voluntary programs, it may not be appropriate to infer the impact on the entire eligible population from the impact on participants when random assignment is conducted as shown in Exhibit 1. That approach assumes that the participation rate among eligibles obtained in the experiment accurately represents the participation rate that would occur in an ongoing program. Suppose instead that we would expect virtually universal participation in an ongoing program, but that only a subset

of eligibles would respond to outreach in a demonstration. For example, in the long run one would expect virtually all eligibles to participate in a publicly funded health insurance program, but only a subset to participate in a demonstration program. Indeed, if outreach is conducted through general media advertising, it may not even be possible to compute a participation rate, since we would not know how many eligible persons were exposed to outreach, and therefore had the opportunity to participate. In such cases, then, a different sample intake strategy is indicated, even for voluntary programs.

Exhibit 2 shows the appropriate intake process for an experiment designed to estimate the impacts of a voluntary program on the eligible population. Exhibit 3 shows the corresponding process for a mandatory program.

In both cases, the process begins with identification of a sample of eligible individuals within the general population. How this is done depends on the nature of the population. In the income maintenance experiments, for example, screening interviews were conducted with a random sample of households in the experimental sites; households with children and income below a specified

level (adjusted for family size) were deemed eligible for the program. In mandatory work programs for welfare recipients, the eligible population within the existing caseload can usually be determined from data maintained in case records; eligibles among the new applicants to welfare can be identified as part of the regular welfare application process.

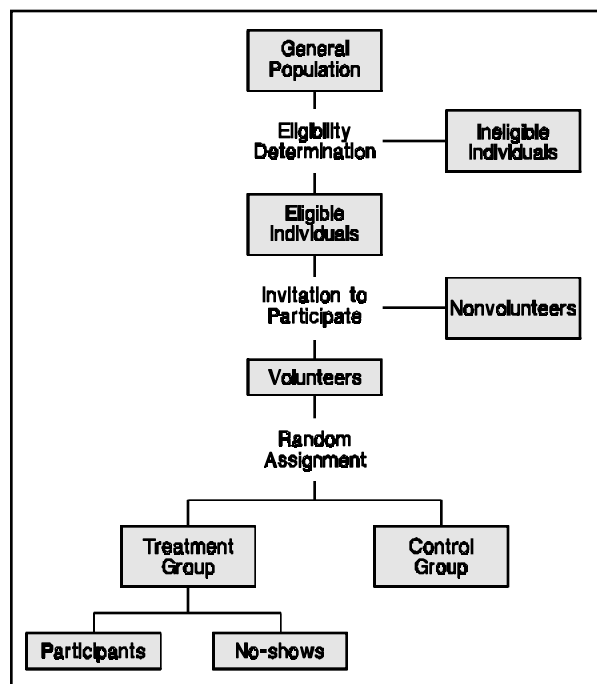
In voluntary programs, those identified as eligible are then invited to participate in the program. At this stage, the information about the program and the experimental evaluation described in the previous section is conveyed to the potential applicants. Because eligibility for the program was determined at the outset, there is no need for an application process (although baseline data may be collected at this point, using a form similar to an application form).² All those who agree to participate (the “volunteers” in Exhibit 2) are randomly assigned to treatment or control status.³

If all those invited to participate agree to do so, the population randomly assigned—and therefore the experimental treatment and control groups—will accurately represent the eligible population.⁴ Of course, that ideal is seldom attained, because some of those invited to participate refuse. It will, however, at least be possible to describe the differences between the experimental sample and the eligible population, using baseline data collected prior to the offer to participate.

In mandatory programs, it is important to include in the treatment group those who refuse to participate in the program, since the program may affect their outcomes.

Design for Estimating Impacts on the Population Eligible for a Voluntary Program

EXHIBIT 2

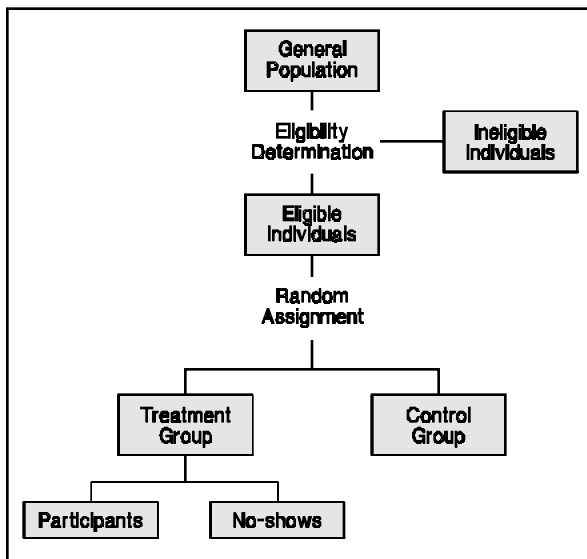


² In subsequent papers, we will discuss the uses of baseline data and procedures for its collection.

³ Alternatively, all eligibles could be randomly assigned and the invitation to participate extended only to the treatment group. Those who refuse the invitation could then be treated as no-shows and the impact on participants derived from the estimated impact on all eligibles using the no-show correction. Since only the treatment group is invited to participate, this approach reduces the number of individuals to whom the program must be explained, avoids the necessity of explaining random assignment to potential participants, and eliminates the necessity of informing controls that they will not be allowed to participate. However, for any given sample size, impacts on participants are estimated less precisely if all eligibles are randomly assigned than if only those who agree to participate are randomly assigned. (We will discuss the effect of the position of random assignment on the precision of the estimates in a subsequent paper.)

⁴ This statement, and similar statements below, apply to the eligible population within the experimental sites—i.e., the eligible population identified in the first step of the intake process. Whether this population is representative of the broader eligible population in other localities, and the implications if it is not, are issues we will address in a subsequent paper.

Design for Estimating Impacts on the Population Eligible for a Mandatory Program

EXHIBIT 3


Therefore, all those identified as eligible are randomly assigned. Those assigned to the treatment group are then informed of the program requirements (see Exhibit 3). There is no need to contact the control group at all, other than to collect any baseline and follow-up data that may be needed. In this case, the experimental treatment and control groups accurately reflect the eligible population and the experimental estimates will be unbiased estimates of program impact on that population.⁵

Evaluations in the Context of Ongoing Programs

Two types of experiment are carried out in the context of ongoing programs: those in which the ongoing program itself is being evaluated and those in which a new treatment for participants in the ongoing program is being tested. An example of the latter is a test of a new employment and training program (the new treatment) for recipients of AFDC (the ongoing program). The design of random assignment in evaluations in the context of ongoing programs will

closely resemble the designs described above for special demonstrations. However, certain constraints apply to experiments undertaken within ongoing programs that are not present in special demonstrations.

Perhaps the most important of these constraints is that the experimenter must work within the established program intake process. In a special demonstration, the intake process can be designed to facilitate the evaluation. In an ongoing program, changing the intake process risks changing the composition of the participant population. Even so simple a change as requiring an additional visit to the program office to meet with intake staff could serve to screen out individuals with transportation or child care problems. Moreover, the experimenter generally has little control over program staff and therefore often cannot convince them to change their established ways of doing things even if it would substantially improve the evaluation.

While the intake processes described above are quite general and apply to most ongoing programs as well as to special demonstrations, variations in the specific ways that the intake staff of existing programs interact with potential participants may make it difficult to inform potential applicants about the evaluation and perform random assignment in the manner that would be optimal from an evaluation standpoint.

For example, in an experimental evaluation of the California Conservation Corps (CCC), the population to be randomly assigned included potential participants in a residential program in one district of the state. Program outreach for the residential component of CCC is conducted by recruiters throughout the state and potential participants are assigned to a particular district partly on the basis of the recruiter's recommendation. In this programmatic context, recruiters had an incentive to recommend that their applicants be assigned to districts other than the one in which the experiment was located, in order to avoid having them assigned to the control group. Only by conducting random assignment after potential participants had been assigned to districts *and doing so without the recruiters' prior knowledge* could the experiment be designed to eliminate the possibility of recruiters "gaming" the random assignment process in this way. This approach to random assignment was deemed both unacceptable and infeasible. Instead, the experimenters met with the recruiters, explained the experimental objectives and procedures, and obtained their agreement to maintain the same pattern of recommendations they would have made in the absence of the experiment. Program staff at the state level then monitored geographic assignments to ensure an adequate flow of potential participants to the experimental district. While

⁵ This assumes that the program being tested would not affect the composition of the eligible population if it were implemented on an ongoing basis. This may not be the case. A mandatory work program for welfare recipients, for example, might deter some individuals from applying for welfare. Since this effect cannot be captured experimentally, the experimental sample would not perfectly represent the population that would be eligible for an ongoing program.

there is no way to verify that the recruiters' recommendations were not influenced by the experiment, this was probably the best compromise that could be obtained in this programmatic context.

Designs to Address Alternative Policy Questions

Up to this point, we have discussed random assignment models in which a single treatment group is compared with a single control group. Such designs address a relatively simple—though often important—factual question: Does the program in question achieve its objectives? The answer to this question bears on an equally simple policy question: *Should the experimental program be adopted or (in the case of an ongoing program) continued?*⁶ More complex policy questions require more complex experimental designs. In this section, we discuss the designs required to address several other policy questions.

First, policy decisions frequently focus on *choices among alternative program strategies*, rather than simply whether or not to adopt (or continue) a particular program. In such situations, it is most useful to test several alternative program approaches in a single experiment, to determine which is the most effective. When the alternatives being compared can be characterized as points along a continuum, the experiment can be designed to *estimate underlying behavioral relationships*. For example, the Health Insurance Experiment estimated the demand for medical care as a function of the net price to the consumer by providing a range of experimental insurance plans with different cost-sharing provisions.

A second type of policy question arises when the program of interest has multiple components and there is interest in the *separate effects of the individual components*. For example, a job training program may provide a number of different employment and training services; in deciding on the optimal mix of services, it is useful to know the impact of each on program participants. In this case, the design of the experiment will depend on whether the program of interest is a new program or an ongoing program.

In this section, we discuss the design of these more complex experiments, presenting detailed examples of each. While these designs are similar in that each involves ran-

dom assignment to multiple treatment groups, they are designed to answer very different policy questions. Therefore, it is essential that evaluators and sponsoring agencies be very clear about the policy question to be addressed before committing to a specific design.

Choosing Among Alternative Program Strategies

Exhibit 4 shows the intake and random assignment process for an experimental demonstration designed to compare two alternative program strategies. (For simplicity of exposition, the stages of outreach and intake leading up to selection of potential participants have been compressed into a single box.) In this design, potential participants are randomly assigned to three different groups: Treatment A, Treatment B, and a control group.

The impact of Treatment A is estimated by the difference in outcomes between those assigned to that treatment and the control group. Similarly, the impact of Treatment B is estimated by the difference in outcomes between those assigned to Treatment B and the control group. This design is thus essentially two experiments in one.

It might seem that one could dispense with the control group and simply compare the outcomes of those assigned to Treatment A with the outcomes of those assigned to Treatment B. That approach is not advisable, for two reasons. First, under that approach one might learn whether Treatment A is preferred to Treatment B, but one would not learn whether either of these strategies is superior to the existing policy regime. Only by including a control group representing the *status quo* can one determine whether either of the new program strategies are a cost-effective improvement over existing programs.

A second reason for including a control group is that, without one, it may not even be possible to obtain a valid comparison of outcomes under Treatment A and those under Treatment B. This will be the case if policy interest focuses on impacts on participants and a nonnegligible number of treatment group members do not participate. Without a control group, impacts on participants cannot be estimated.⁷

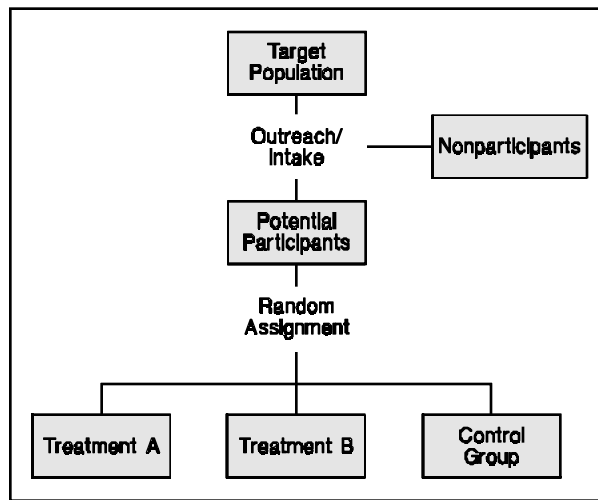
An important feature of this design is that random assignment creates three groups that do not differ systematically except for experimental treatment. Thus, the experimental

⁶ As we will see in a subsequent paper, the experimental impact estimates provide information that is necessary, but not sufficient, to address this question. In addition, one must take into account the costs of the program.

⁷ If the participation rate were the same in both treatments, the no-show correction would be the same for the two, so that the impacts of the two experimental programs on participants would be proportional to their impacts on the overall treatment group. One cannot know in advance, however, what the participation rates in the two programs will be.

Design for Estimating Impacts of Alternative Programs

EXHIBIT 4



estimates of the impacts of Treatments A and B pertain to the *same population*, the potential participants who were randomly assigned. This means that the two impact estimates can be directly compared and the more effective treatment for this population determined. (As we will see, this is not true of all designs in which experimental impact estimates are derived for multiple treatments.) Estimates of the impact of different treatments on the same population are called **differential impact estimates**.

Illustrative Example:

The Moving to Opportunity Demonstration.

The federal government has traditionally provided two types of housing assistance for low-income families: public housing and rent vouchers. Public housing units are owned by local government agencies and rented to low-income families at below-market rates. Rent vouchers subsidize a portion of the cost of housing units rented by low-income families from landlords in the private housing market. Rent vouchers therefore avoid the concentration of large numbers of poor families in a small area that is typical of public housing projects. Nevertheless, families receiving rent vouchers tend to locate in high-poverty neighborhoods, in part because of their family and social ties to those areas. The endemic social problems and limited employment and educational opportunities in these neighborhoods arguably perpetuate a cycle of poverty from one generation to the next.

The Moving to Opportunity (MTO) Demonstration is designed to test one strategy for breaking this cycle of poverty: housing assistance that encourages poor families to move

from high-poverty areas to low-poverty areas. The specific intervention used in the demonstration is rent vouchers that can be used only in low-poverty areas, coupled with intensive counseling and assistance in finding suitable, affordable units in those areas. This approach is patterned on programs in Chicago and other cities that were ordered by the courts as a way to reduce racial segregation. Nonexperimental evaluations of these programs had indicated substantial gains in earnings and educational attainment among participating families who moved from high-poverty to low-poverty areas.⁸ In 1994, the U.S. Department of Housing and Urban Development (HUD) initiated an experimental demonstration to test whether those gains were really attributable to the differences between the two environments, rather than to selection effects among the families who chose to move to low-poverty areas. Geographically restricted vouchers were chosen as the experimental intervention, not because there was policy interest in such vouchers per se, but as a way to create comparable groups of families in the two environments. If the experiment demonstrates that a low-poverty environment has positive effects on the well-being of poor families, policies can then be devised to encourage dispersion of poor families to such areas.

In the MTO demonstration, families living in public housing projects are invited to apply for vouchers that allow them to move to private housing. Potential participants are randomly assigned to one of three groups:

- the **MTO voucher group**, which receives rent vouchers that can be used only in low-poverty areas, along with intensive counseling and assistance in finding suitable housing in those areas;
- the **regular voucher group**, which receives traditional rent vouchers that can be used anywhere in the area;
- a **control group**, which receives no vouchers, but is allowed to continue to live in public housing.

Demonstration intake is being conducted by local public housing authorities in five large cities (Baltimore, Boston, Chicago, Los Angeles, and New York). Counseling and relocation assistance are being provided by local nonprofit organizations, to help families find private housing that meets minimum quality standards, with a landlord who is willing to accept the voucher. If they are unsuccessful in finding housing that meets these requirements, they are eligible to remain in public housing. Families in the experiment are guaranteed housing subsidies for 5 years.

⁸ See Rosenbaum (1991).

The design calls for a total of approximately 4,400 families to be assigned to the three experimental groups.⁹ Follow-up interviews with these families will be conducted over a 10-year period, to allow estimation of long-term impacts on their employment, income, education, and social well-being. Long-term effects on the children in participating families will be of particular interest.

The impact analysis will compare the effects of the two different environments *relative to public housing*. That is, the difference in outcomes between the MTO voucher group and the control group provides an estimate of the effects of a low-poverty environment, relative to living in public housing in a high-poverty environment. Similarly, the difference in outcomes between those receiving traditional (unrestricted) vouchers and the control group is an estimate of the net effects of living in private housing in a high-poverty area, relative to living in public housing in the same type of area. These two impact estimates can be compared to determine the relative effects of living in low-poverty vs. high-poverty areas, against the common counterfactual of living in public housing.

Since the experimental design involved randomly assigning a common pool of potential participants to the three treatments, the groups assigned to MTO vouchers and regular vouchers are well-matched; thus, the impacts of these two treatments *on the entire treatment group* are directly comparable. That is, we will be able to say which approach had the larger effects (relative to public housing) on the entire group that was randomly assigned and given vouchers.

Not all of the families in the two treatment groups that receive vouchers will be successful in finding private housing, however. Since the policy interest is in the relative effects of the two different environments, we would really like to compare the impacts on those who actually move to low-poverty areas with those on families who remain in high-poverty areas. If we are willing to assume that the experiment has no impact on families who are unable to use the vouchers to obtain private housing, we can use the no-show correction discussed in the previous paper to estimate the impacts on the subgroups who are successful in using the vouchers to rent private housing and move out of public housing. Unfortunately, these subgroups are *not* necessarily comparable, because the two voucher groups have different success rates: in the MTO voucher group, the success rate is running 60-70 percent, whereas in the regu-

lar voucher group, it is 80-90 percent. Therefore, in comparing the estimates of impact on successful families, it will be necessary to take into account any differences in the characteristics of these two populations.¹⁰

To the extent that participating families are successful in moving to, and remaining in, low-poverty areas, the MTO demonstration will provide for the first time reliable measures of the effects of the social and economic environment on the well-being of poor families and their children. It will compare the experiences of families living in poverty-stricken neighborhoods with those of families in the much richer social, economic, and educational environments of low-poverty areas. Only through an experimental design can one generate samples of comparable families living in these different environments, in order to measure the effects of the environments themselves. Knowledge of those effects can be invaluable in assessing a wide range of public policies.

Estimating Behavioral Response Functions

A particularly powerful experimental design can be implemented when the program variants of interest can be characterized as points along a policy continuum. In such cases, one can estimate a **behavioral response function**, which shows how the effects of the program will change as its parameters change. For example, in the Housing Allowance Demand Experiment, families in different experimental groups received housing subsidies equal to 0, 20, 30, 40, 50, or 60 percent of their rent.¹¹ The income maintenance experiments estimated the labor supply response of low-income families to a wide range of cash transfer programs that differed in the benefit provided a family with no other income (the “guarantee”) and the rate at which that benefit was reduced as the family’s earnings rose (the “tax rate”).

Where feasible, estimation of behavioral response functions can provide policymakers with an extremely valuable tool for policy analysis. Not only do such functions allow

⁹ Additional families may be added in several sites if, as anticipated, additional vouchers are made available to the demonstration by the local public housing authorities.

¹⁰ To do so formally will require application of nonexperimental statistical adjustments of the impact estimates. Thus, the research question cannot be fully addressed with experimental methods. Nevertheless, random assignment is extremely valuable in this case because it avoids the potentially severe selection bias that would be involved in comparing families who were successful in moving to low-poverty areas with a comparison group composed entirely of families who were unsuccessful or, worse yet, who were not interested in moving.

¹¹ See Friedman and Weinberg (1983). The Housing Allowance Demand Experiment also tested other, more complex, subsidy formulae.

the analyst to interpolate responses to parameter values between those tested, or extrapolate beyond the range of values tested; the fundamental behavioral relationships underlying such functions may also be applicable to policies that are entirely different from those originally tested. The labor supply parameters estimated in the income maintenance experiments were used to analyze not only a variety of negative income tax plans in the 1970s, but also other policies that involved cash payments to low-income families with significant tax rates on earnings, such as the Earned Income Tax Credit. Similarly, the estimates of the price-sensitivity of demand for medical care derived from the Health Insurance Experiment and the demand for housing from the Housing Allowance Demand Experiment are applicable to a wide range of policy analyses that involve consumption subsidies in those markets.

Unfortunately, the range of programs and policies that lend themselves to estimation of behavioral response functions may be fairly limited. The central features of most programs simply cannot be expressed as quantitative parameters that can be varied continuously. In a job training program, for example, the critical design features include the type of training provided (*e.g.*, classroom training vs. on-the-job training), the content of the curriculum and occupation for which participants are trained, the skills and qualifications of the training staff, the nature of linkages to private employers, and other nonquantitative program characteristics. In the case of such programs, the best one can do is to experiment with alternative combinations of these central features—*e.g.*, classroom training with and without close linkages to private employers and/or on-the-job training in alternative occupations. (We discuss below the design of experiments to estimate the effects of discrete program components.)

Illustrative Example: The Health Insurance Experiment.¹²

The Health Insurance Experiment grew out of the debate in the early 1970s over proposals to provide universal health insurance, either through mandated employment-related coverage or direct government provision.¹³ Efforts

to estimate the cost of these plans prompted a spirited debate about the effects that increased insurance coverage would have on the use of medical care. Economists pointed out that extension of health insurance to previously uncovered individuals constituted a substantial reduction in the net price of medical care to those individuals and could therefore be expected to result in an increase in the demand for care. To offset this increase in demand, which could be extremely expensive and/or inflationary, some analysts proposed that any such plan include “cost-sharing” in the form of deductibles or coinsurance (an initial amount or a percentage of the bill to be paid by the beneficiary). Others argued that the use of medical care was not sensitive to monetary prices (Fein, 1971) or that cost-sharing provisions would deter individuals, especially the poor, from receiving needed care.

Efforts to estimate the price elasticity of demand for medical care nonexperimentally were hampered by lack of adequate data and faced a serious threat of selection bias. Individuals who expect to incur high medical expenses have a greater incentive to purchase health insurance than those who do not. This means that those with the highest expenditures may face the lowest prices (*i.e.*, have the best insurance). Such “adverse selection” would create an upward bias in nonexperimental estimates of the effect of *changes* in the net price of medical care on consumption of care.

The Health Insurance Experiment estimated the price elasticity of demand for medical care by randomly assigning families to insurance plans with different cost-sharing provisions. Under the experimental plans, the family either received full reimbursement of all medical costs (the “free plan”) or was required to pay 25, 50, or 95 percent of the cost of covered services, up to an annual limit that varied with family income, but was capped at \$1,000. Above this limit, the plan paid all medical costs.¹⁴ The experimental policies covered a comprehensive range of inpatient and outpatient medical, dental, and mental health services.

A notable feature of the experiment was that it had no control group; all analyses were based on comparisons among

¹² For a detailed description of the Health Insurance Experiment, see Newhouse (1993).

¹³ A subsidiary issue that played an important role in the initiation of the experiment was related to the work disincentive posed by the Medicaid “notch”—the abrupt cessation of all benefits when covered families’ earnings rose above the Medicaid eligibility level. Some policy analysts proposed smoothly phasing out health insurance coverage of the poor through the use of income-related cost-sharing under which the share paid by the beneficiary would rise with income. Feldstein (1971) proposed this approach as a way to ensure that cost-sharing did not deter the poor from receiving needed care under a universal plan with cost-sharing.

¹⁴ The experiment included 15 treatments. Ten comprised all the possible combinations of the 25, 50, and 95 percent coinsurance rates with three different levels of the annual expenditure limit (5, 10, and 15 percent of family income, up to \$1,000), plus the free (0 percent coinsurance) plan, to which the annual limit did not apply, because the plan paid all the family’s medical expenses. Four of the remaining experimental treatments incorporated different cost-sharing provisions for different types of care (inpatient vs. outpatient; mental health and dental care vs. all other services). The final treatment was enrollment in a Health Maintenance Organization (HMO). In addition, data were collected on a representative sample of regular enrollees in the same HMO, for comparison with the randomly assigned HMO sample.

the experimental plans. This reflects its fundamental objective of estimating the behavioral response to experimental variation in the price of care rather than estimating the impact of changing that price from the status quo to a different level.

Reflecting the policy interest in a national health insurance program with universal coverage, the experimental sample was drawn to represent the general population under the age of 65. The aged were excluded on the grounds that, because they were already covered by Medicare, they were unlikely to be strongly affected by any new national health insurance plan and because their behavior was likely to be sufficiently different from that of the nonaged population that they would have required a separate experiment.

A random sample of the nonaged population in six sites was identified through screening surveys, assigned to the experimental treatments, and then invited to enroll in the experiment. Those who agreed to participate received coverage under the experimental insurance plans for either three years or five years.

The difference in duration of coverage was one of several “subexperiments” conducted within the larger experiment. A randomly selected 25 percent of the sample in each insurance plan (50 percent in the first site) was assigned to receive coverage for five years. This allowed better estimation of steady-state impacts on demand for care and health status, and provided a test for any bias that might have arisen from the limited duration of the experiment. Other subexperiments measured the effect of cash participation incentive payments, mail questionnaires, and initial physical examinations on the utilization of care.

Over the period 1974-77, nearly 6,000 individuals were enrolled in the experiment.¹⁷ The experimental plans were administered by a commercial claims processing firm, under contract to the research organization that designed the experiment.

Data on medical care costs and utilization were derived from the insurance claims submitted to the experiment. Baseline and follow-up data on physical, mental, and social health were collected through a combination of personal interviews and mail questionnaires; in addition, physical examinations were administered at baseline and at the end of the enrollment period.

The experimental results on the central issue of the price elasticity of demand were clear and striking. Overall medical expenditures were 45 percent higher under the free plan than under the 95 percent coinsurance plan; outpatient expenditures were 68 percent higher, while inpatient costs were 30 percent higher. Exhibit 5 shows the response function for total annual expenditures per participant, estimated from the experimental data. The findings with respect to health status were equally striking: On the wide range of outcomes measured, the additional care induced by the free plan had little or no beneficial effect on health status. The clear policy implication was that cost-sharing is an effective way to contain health care costs and utilization, and that doing so will not have a deleterious effect on the health of covered individuals.

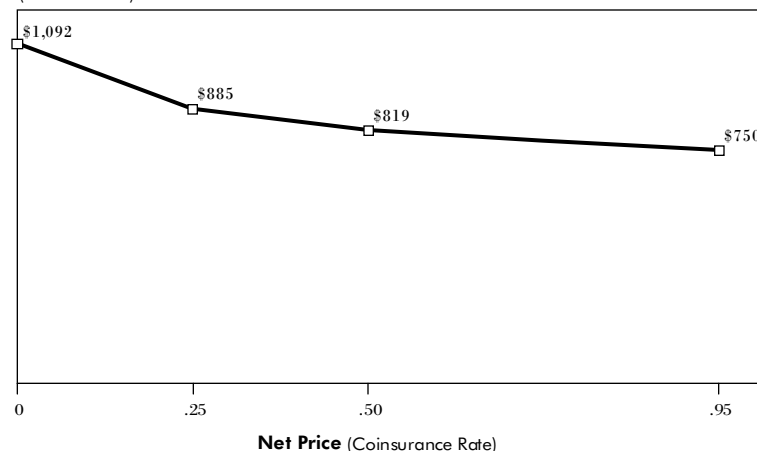
¹⁷ Of these, approximately 4,000 were enrolled in fee-for-service plans and about 1,800 were enrolled in a Health Maintenance Organization (HMO).

Demand for Medical Care — Estimates from the Health Insurance Experiment

EXHIBIT 5

Annual Medical Expenditures per Participant

(1993 Dollars)



When the Health Insurance Experiment was begun in the early 1970s, the national policy debate concerned expanding access to medical care through a national health insurance program. By the time the experimental results became available in the 1980s, concern had switched to cost containment and limiting overutilization, although the early 1990s saw at least a brief revival of interest in expanded coverage. By focusing on the fundamental issue of consumer response to the price incentives embodied in health insurance, the experiment was able to provide results that were highly relevant in both policy environments.¹⁸

Estimating the Effects of Discrete Program Components in Special Demonstrations

A third situation that calls for multiple treatment groups arises when the program of interest has multiple components and there is policy interest in their separate effects. This would be the case, for example, if policymakers were trying to decide what combination of provisions to include in a new program. In contrast to the case discussed in the previous section, where the different treatments were viewed as *alternatives*, in this case the treatments are being considered for use *in combination*. The complication presented by this situation is that the effects of a given component may vary, depending on the other components in the package. Thus, one must include in the experiment not only all the components, but all the feasible *combinations* of components.

Consider, for example, a welfare reform proposal consisting of employment and training services for current recipients and a guarantee of child care for a year after leaving the welfare rolls to take a job. Proponents of such a proposal might argue that employment and training services are ineffective in the absence of child care and that child care alone will not help recipients become employed, but that in combination the two can help recipients obtain and hold jobs. Others might contend that employment and training services alone would be sufficient and that, once they are provided, the child care guarantee would add little but additional cost. Still others might argue the reverse, that child care is the binding constraint and that once it is provided employment and training services are unnecessary. To determine which of these conflicting positions is correct one must estimate not only the impacts of the overall proposal, but also those of the two separate components, taken by themselves.

Factorial Design for Estimating Impacts of Discrete Program Components

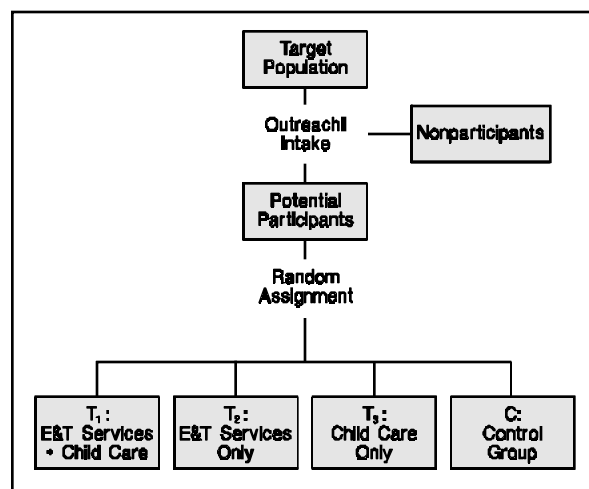
EXHIBIT 6


Exhibit 6 shows the experimental design required to produce these estimates. In this design, potential participants are randomly assigned to four groups:

- T₁**: a treatment group that receives both employment and training services and the child care guarantee;
- T₂**: a treatment group that receives only employment and training services;
- T₃**: a treatment group that receives only the child care guarantee; and,
- C**: a control group that receives no additional services.

Comparison of the outcomes of each of the treatment groups with those of the control group provides experimental estimates of the impact of each combination of services. As before, these are *differential impact estimates*, which show the impact of different policies on the same population. Thus, the impact estimates allow us to choose the most effective combination of program components for that population.

This design is an example of a **factorial** or **fully randomized** experimental design, in which all possible combinations of two or more treatments are tested.¹⁹ The need to test all possible combinations arises from the possibility of **interactions** between the treatments—*i.e.*, the possibility that the combined effect of the two treatments is different from the sum of their individual effects. If there were no interactions, it would be possible to compute the

¹⁸ The experimental results for the HMO plan, not discussed here, are also highly relevant to the recent policy interest in managed care plans.

¹⁹ See Campbell and Stanley (1963) for a detailed discussion of factorial designs.

effect of the combined treatment (T_1) by simply summing the effects of the two individual treatments (T_2 and T_3). The treatment group receiving both treatments would be unnecessary; this would allow the experiment to be conducted more cheaply or, alternatively, it would allow larger sample sizes in the experimental groups, thereby increasing the statistical precision of the estimates. Such efficiencies come at a cost, however: in adopting a design that assumes no interactions, we run the risk of seriously misestimating the combined effect of the two treatments if there really are interactions between the program components.

A better strategy is to adopt the complete factorial design and then, in the analysis, test for the presence of interactions. If the test shows that interactions are negligible, treatment T_1 can be combined with treatment T_2 in estimating the effects of the employment and training component and with treatment T_3 in estimating the effects of the child care component, thereby recouping the apparent loss of sample size in each treatment in the factorial design.²⁰

Illustrative Example: The New Jersey Income Maintenance Experiment.

Factorial designs arise quite naturally in the design of experiments to estimate behavioral response functions when the policy of interest is characterized by two or more parameters. The income maintenance experiments, for example, were designed to test the labor supply response to cash transfers in the form of a negative income tax. In its simplest form, the negative income tax is defined by two parameters: the **guarantee**, which is the amount of the transfer to a family with no other income, and the **tax rate**, which is the rate at which the transfer is reduced for each dollar of earnings. Because each of these parameters can be varied independently over a wide range, there was strong interest in learning the labor supply response to variations in each. Moreover, there were theoretical reasons to expect the labor supply effects of a given guarantee to depend on the level of the tax rate, and *vice versa*—i.e., interaction effects were expected. Therefore, factorial designs were adopted in all of the income maintenance experiments.

The New Jersey Income Maintenance Experiment, for example, was originally designed to test three tax rates (30, 50, and 70 percent) and three guarantee levels (50, 75, and 100 percent of the poverty line). As shown in Exhibit

Experimental Treatments—New Jersey Income Maintenance Experiment

EXHIBIT 7

Guarantee (percent of poverty level)	TAX RATE (Percent)		
	30	50	70
50%	X	X	
75%	X	X	X
100%		X	X

Source: Kershaw and Fair (1976), p.9.

Note: This is the original design of the New Jersey Experiment; an eighth experimental treatment with a guarantee at 125 percent of the poverty level and a 50 percent tax rate was subsequently added.

7, only seven of the nine possible combinations of these parameter values (those indicated by Xs) were included in the design, however.²¹ The combination of the highest guarantee and the lowest tax rate was deemed too generous, and the combination of the lowest guarantee and highest tax rate not generous enough, to be relevant for policy. Thus, the design is an **incomplete factorial design**.

The experiment was fielded in four cities in New Jersey and one in Pennsylvania over the period 1968-72.²² Screening surveys were conducted in low-income areas in those cities to identify families headed by nonaged males with incomes below 150 percent of the poverty line; these families were then randomly assigned to treatment and control groups and the treatment group members were invited to enroll in the experiment. Thus, the experimental sample was designed to be representative of the entire population of low-income, nonaged two-parent families in those areas.

During their three-year enrollment period, the 725 families in the treatment groups filed monthly reports on their earnings and other income, on which their monthly negative income tax payments were based. The payments were administered by the research organizations running the experiment, according to rules and procedures designed

²⁰ Doing so requires a multivariate regression analysis, in which the effects of both components are estimated simultaneously, rather than a simple comparison of mean outcomes. We will discuss this analytic approach in a subsequent paper.

²¹ An eighth plan, incorporating a fourth guarantee level (125 percent of the poverty level) and a 50 percent tax rate, was subsequently adopted when New Jersey adopted a welfare program for two-parent families headed by unemployed workers that provided more generous benefits than several of the experimental plans. See Kershaw and Fair (1976) and Watts and Rees (1977) for a detailed discussion of the design of the New Jersey Experiment.

²² A companion project begun a year later, the Rural Income Maintenance Experiment, tested similar treatments in rural areas.

especially for the experiment. The 632 control families received no NIT payments, but remained eligible for any other publicly provided payments or benefits.²³ Follow-up data on employment and earnings, expenditures, family composition, and a variety of other social, economic, and attitudinal outcomes were collected through personal interviews with family members.

Among male heads of family, the experiment found small reductions in labor supply that were not statistically significantly different from zero. Wives' labor supply responses were much larger, with reductions on the order of 20 percent. No consistent, statistically significant differences in impact on work effort were found among the NIT plans. This probably reflects the small samples enrolled in the different plans, the small overall response of male heads of family, and the small number of working wives in the sample. The most positive finding was that the experimental plans increased high school completion rates among children in the treatment group families by 25 to 50 percent.

The New Jersey Experiment was designed primarily to test whether extension of cash transfers to intact families (who had traditionally been excluded from AFDC) would cause them to work less, an issue that was central to academic discussions of the negative income tax and that was expected to be important in the national policy debate on welfare reform. The experimental results were relatively reassuring on that question; together with the results of three other similar income maintenance experiments, they probably played an important role in neutralizing the work effort issue in the national policy debate. However, throughout the 1970s all attempts to legislate such an extension were defeated by an unusual alliance of conservatives who opposed extension of cash transfers to intact families on cost and equity grounds and liberals who viewed all politically viable plans as providing inadequate benefits. In 1981, the focus shifted to reducing the existing welfare rolls, and interest in cash transfers for intact families disappeared.

Although one cannot attribute any specific policy decision to the findings of the New Jersey Experiment, it and the other income maintenance experiments that were patterned after it provided valuable information on the labor market behavior of low-income families that has been used extensively in the analysis of a wide range of policy options. In

particular, labor supply elasticities based on the larger samples available from the Seattle-Denver Income Maintenance Experiment were built into the simulation model that was used by the Department of Health, Education, and Welfare (now the Department of Health and Human Services) to estimate the costs and distributional consequences of virtually all the welfare reform proposals of the 1970s. And, as the first highly visible large-scale social experiment, the New Jersey Experiment had an enormous influence on the development and widespread use of experimental methods for the evaluation of social policies.

Estimating the Effects of Discrete Program Components in Ongoing Programs

The designs discussed up to this point are useful when policymakers want to compare alternative policy options, whether they be alternative program strategies or alternative combinations of program components. In these designs, a common pool of potential participants are randomly assigned to one or more policy options and a control group. This creates well-matched treatment and control groups and allows estimation of the effects of different policies on the same population, *i.e.*, differential impact estimates. In certain circumstances, however, it is useful to know the effects of different policies on *different* populations.

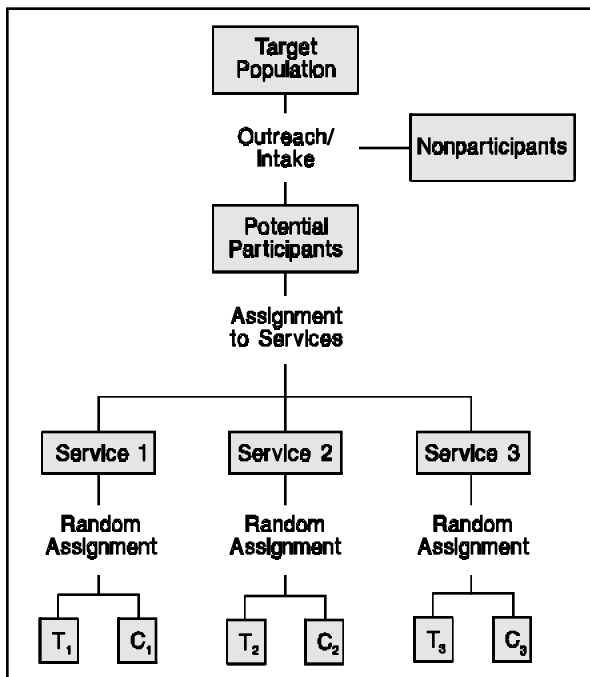
This will be the case when evaluating an ongoing program that has several distinct components that are applied to different participants, either at the discretion of program staff or by choice of the participants. For example, job training programs typically provide a range of different services, such as job search assistance, occupational skills training, and on-the-job training. The specific services to be received by each participant are determined through an ongoing interaction between the participant and program staff once the participant is enrolled in the program. Some participants receive multiple services, either simultaneously or sequentially, and service plans may change over time, depending on the results of initial services.

In this case, direct random assignment of potential participants to different program services would run counter to the fundamental objective of the evaluation, which is to measure the impacts of the program *as it normally operates*. Part of the normal operation of the program is the application of staff judgment or participant preferences in deciding which participants are to receive specific services. Overriding these judgments or preferences with random assignment to services would certainly change the way the program operates and would arguably reduce its effectiveness by providing inappropriate services to some participants.

²³ When the experiment began, the principal welfare program in New Jersey and Pennsylvania, Aid to Families with Dependent Children (AFDC), excluded families with two able-bodied parents. Thus, virtually all of the control families were ineligible for welfare at the outset. In January 1969, however, New Jersey instituted an AFDC program for two-parent families (AFDC-UP) with benefits that were among the highest in the country.

Design for Estimating Impacts of Discrete Program Components in an Ongoing Program

EXHIBIT 8



Ideally, one would like to conduct random assignment after program staff have assigned potential participants to services, as shown in Exhibit 8. This design provides a well-matched control group for the participants assigned to each service. Thus, the impact of each service on those assigned to it can be estimated. In effect, this design provides a separate experiment for each program component. In addition, an estimate of the overall impact of the program on the entire participant population can be obtained by comparing the entire treatment group, across all components, with the entire control group.

It is important to recognize, however, that because assignment to service was judgmental, the groups assigned to different services are potentially different. This means that the estimated impacts of different services are not directly comparable. Suppose, for example, that service A is found to have a larger positive impact on those assigned to it than service B has on those assigned to it. We cannot therefore conclude that service A is “better” in the sense that if those assigned to service B had instead been assigned to service A they would have experienced larger impacts. It may be that those assigned to service A were more motivated or talented than those assigned to service B. Or it may be that even though service A works well for those assigned to it, it would not be a good match to the abilities and aptitudes of those assigned to service B. All we can say from this result is that service A works better for those assigned to it than service B does for those assigned to it.

Such a result is nevertheless of substantial value to policymakers. It allows them to identify the components of the program that are working well for those they serve and those components that are not. This in turn allows policymakers to focus their attention on those components in need of improvement or elimination, rather than continuing to spend resources on ineffective services. But it does not tell policymakers *how* to improve those components; to do that would require testing alternatives to the ineffective components *for the populations that they serve*.

This means that improving existing programs must be a two-stage process. First, one must determine which parts of the program are achieving their objectives and which are not. For this purpose, experimental designs of the type described in this section, with assignment to program components prior to random assignment, are appropriate. Then, one must test alternatives to those program components found to be ineffective in the first stage, using designs like those described in the previous section, which randomly assign potential participants to different program components, thereby yielding differential impact estimates.

Illustrative Example: The National JTPA Study.

Since 1962, the federal government has provided job training for economically disadvantaged workers. Over that period, a variety of attempts have been made to estimate the impact of such training on participants’ earnings. One of the most ambitious attempts to evaluate these programs was a series of comparison group evaluations of the Comprehensive Employment and Training Act (CETA) program in the late 1970s and early 1980s (see discussion in the first paper in this series). These evaluations revealed that, even when the same data base was employed, different nonexperimental methods gave substantially different impact estimates.²⁴ This result was confirmed in methodological studies that applied various different nonexperimental estimation techniques to data from an employment and training demonstration where unbiased experimental estimates of impact were available.²⁵

Because of the uncertainties involved in nonexperimental evaluation methods, a panel of experts convened by the Department of Labor to advise it on the evaluation of the program that replaced CETA, the Job Training Partnership Act (JTPA), unanimously recommended that the evaluation employ experimental methods.²⁶ The resulting evaluation was conducted in 16 local JTPA service delivery areas over the period 1987-92.

²⁴ See Barnow (1987).

²⁵ See LaLonde (1986) and Maynard and Fraker (1987).

²⁶ See Stromsdorfer et al. (1985).

Because the evaluation was designed to estimate the effects of JTPA on those who normally participate in the program in these sites, program staff conducted program outreach, eligibility determination, and applicant screening in the usual manner. The final step in the regular JTPA intake process is an assessment of applicant needs and interests. On the basis of this assessment, the intake worker recommended one of three *service strategies* for each potential participant: a strategy based on on-the-job training (OJT), one based on classroom training in occupational skills, or one based on other less intensive services.²⁷ The potential participants were randomly assigned after this assessment was conducted and the intake workers' service recommendations were recorded. Thus, the experimental design was essentially that shown in Exhibit 8, with the exception that the intake workers' service recommendations were not strictly binding (see discussion below).

Because service recommendations were made before random assignment, they were not affected by experimental assignment. Thus, the treatment group within each service strategy subgroup was well-matched to its control group and constituted a subexperiment that yielded unbiased estimates of impact on those for whom that service strategy was recommended. In addition, the combined treatment and control groups across all three subgroups yielded an experimental estimate of the overall impact of the program on the entire treatment group.

In interpreting the results of such a design, it is important to bear two caveats in mind. First, as noted earlier, the service strategy subgroup estimates are not differential impact estimates; because potential participants were assigned to the three subgroups judgmentally, rather than randomly, they represent different participant populations. It is clear, for example, both from discussions with program staff and from data on participant characteristics and outcomes, that staff tended to assign the more job-ready applicants to the OJT subgroup.²⁸ This means that the impact estimates derived for one subgroup cannot be ap-

plied to another. For example, among adult trainees it was found that the OJT subgroup had larger earnings gains than the classroom training subgroup, but this does not necessarily mean that if those in the classroom training subgroup had been assigned the OJT subgroup instead they would have had larger earnings gains. It may simply mean that the more job-ready applicants assigned to the OJT subgroup were better able to benefit from the program.

The second caveat that must be borne in mind is that the treatment-control differences for each subgroup represent the impact of *recommending* a certain set of services for that subgroup, not the actual *receipt* of those services. Because the process of matching participants to services is an ongoing one in JTPA, not all treatment group members received the services that were recommended for them and some received services that were not recommended for them. This was the unavoidable outcome of two constraints on the research design: the need to define service subgroups prior to random assignment and the need to avoid disturbing the normal operation of the program. This caveat notwithstanding, however, the three subgroups experienced very different patterns of service receipt that were highly correlated with the service recommendations. Thus, the impacts estimated for the three groups can be viewed as the result of meaningfully different service *strategies*.

Over 20,000 potential participants were randomly assigned in the National JTPA Study sites. Baseline data were collected as part of the program intake process and 30 months of follow-up data on employment and earnings, non-JTPA education and training, welfare benefits, and other economic and social outcomes were collected through a combination of personal interviews and administrative records.

Separate analyses were conducted for adult men, adult women, female youths, and male youths.²⁹ The overall effect of the program on earnings (*i.e.*, the average effect across all three service strategies) was statistically significantly positive for the two adult groups, but not significantly different from zero for either of the two youth groups. Indeed, none of the estimated effects for any of the six youth service strategy subgroups were statistically significantly different from zero. When estimated program benefits in the form of increased earnings were compared with program costs, JTPA had positive net social benefits for five of the six adult service strategy groups, but not for any of the six youth service strategy subgroups.

²⁷ In some cases, multiple services were recommended. These sample members were still categorized on the basis of whether OJT or classroom training in occupational skills (as distinct from basic education) was recommended, regardless of the other services that were recommended, on the grounds that these are the most intensive services that JTPA offers. See Orr et al. (1996) for a detailed description of the National JTPA Study.

²⁸ The clearest way in which this selection process can be gauged is by comparing the post-random assignment earnings of controls in the three service strategy subgroups. Among adult women, for example, those in the OJT subgroup averaged approximately \$15,000 in earnings over the 30-month followup period, while those in the classroom training and less intensive services subgroups averaged only about \$11,500 and \$10,250, respectively (see Orr et al., 1996).

²⁹ In JTPA, participants aged 16-21 are classified as youths. The experimental sample included only out-of-school youths.

These findings had a dramatic and immediate effect on the policy deliberations regarding JTPA. As noted in a previous paper, as a direct result of these findings, annual funding for the youth component of JTPA was reduced by over \$500 million, while funding for the adult components remained essentially unchanged. This represented a savings to taxpayers at no loss to the intended beneficiaries of the program, since the training was totally ineffective. Part of the resources that would have been wasted on ineffective services was devoted instead to experimental demonstrations designed to test alternative training strategies for youths, in the hope of identifying more effective service approaches. As this is written, those experiments have just begun.



References

- Barnow, Burt S. 1987. "The Impact of CETA Programs on Earnings: A Review of the Literature." *Journal of Human Resources* 22 (Spring): 157-93.
- Campbell, Donald T., and Julian C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. (1st ed.) Chicago: Rand-McNally.
- Fein, Rashi. 1971. Testimony on Health Care Crisis in America, Hearings before the Subcommittee on Health of the Committee on Labor and Public Welfare, U.S. Senate. February 22 and 23, p. 146.
- Feldstein, Martin S. 1971. "A New Approach to National Health Insurance," *Public Interest* 23:93-105.
- Friedman, Joseph, and Daniel H. Weinberg (eds.). 1983. *Urban Affairs Annual Review*, Vol. 24: The Great Housing Experiment. Beverly Hills: Sage.
- Kershaw, David, and Jerilyn Fair. 1976. *The New Jersey Income-Maintenance Experiment. Volume I: Operations, Surveys, and Administration*. New York: Academic Press.
- LaLonde, Robert J. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *American Economic Review* 76 (September): 604-20.
- Maynard, Rebecca, and Thomas Fraker. 1987. "The Adequacy of Comparison Group Designs for Evaluations of Employment-Related Programs." *Journal of Human Resources* 22 (Spring): 194-227.
- Newhouse, Joseph P. 1993. *Free for All? Lessons from the RAND Health Insurance Experiment*. Cambridge, Mass.: Harvard University Press.
- Orr, Larry L., Howard S. Bloom, Stephen H. Bell, Fred Doolittle, Winston Lin, and George Cave. 1996. *Does Job Training for the Disadvantaged Work? Evidence from the National JTPA Study*. Washington, D.C.: Urban Institute Press.
- Rosenbaum, James E. 1991. "Black Pioneers—Do Their Moves to the Suburbs Increase Economic Opportunity for Mothers and Children?" *Housing Policy Debate* 2:4.
- Stromsdorfer, E., H. Bloom, R. Boruch, M. Borus, J. Gueron, A. Gustman, P. Rossi, F. Scheuren, M. Smith, and F. Stafford. 1985. *Recommendations of the Job Training Longitudinal Survey Research Advisory Panel*. Washington, D.C.: Employment and Training Administration, U.S. Department of Labor.
- Watts, Harold W., and Albert Rees (eds.). 1977. *The New Jersey Income-Maintenance Experiment. Volume II: Labor-Supply Responses*. New York: Academic Press.